Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.20260454463

AG-BiLSTM: A Hybrid Deep Learning Model for Olympic Medal Prediction

Junyao Xu*

Department of Information Science Education, HoHai University, Changzhou 213200, China

**Corresponding author: Junyao Xu.*

Abstract

Time-series forecasting is an important research area in computer science and has been widely applied to complex prediction tasks. Olympic medal prediction is a typical multivariate time-series forecasting problem that supports national sports planning and resource allocation. However, existing methods often struggle to capture nonlinear temporal patterns and long-range dependencies, particularly when historical data are limited. To address these challenges, this study proposes Attention Gated Bidirectional Long Short-Term Memory (AG-BiLSTM), a hybrid deep learning framework that integrates a Bidirectional Long Short-Term Memory (BiLSTM) network, a multi-head self-attention mechanism, and a gated fusion module. BiLSTM is employed to learn bidirectional temporal dependencies from historical Olympic data, while the attention mechanism identifies critical time steps. The gated fusion module further enhances prediction performance by adaptively weighting heterogeneous features. In addition, a country-level temporal splitting strategy is adopted to improve model generalization across countries with different data distributions. Experiments on a self-constructed dataset covering 19 countries demonstrate that AG-BiLSTM achieves an average Root Mean Square Error (RMSE) of 3.74 and consistently outperforms ARIMA, XGBoost, LSTM, BiLSTM, Transformer, and iTransformer models. The proposed model shows superior predictive accuracy across gold, silver, and bronze medal forecasting tasks, with particularly strong performance in gold medal prediction. These results indicate that AG-BiLSTM is an effective and robust approach for Olympic medal prediction under small-sample conditions.

Keywords

Olympic medal prediction, time series forecasting, AG-BiLSTM, multi-head self-attention, gated fusion

1. Introduction

Forecasting time-series data stands as a key research direction in computer science [1], with wide applications in finance [2], transportation [3], energy systems, and sports analytics. In recent years, sports performance prediction has attracted increasing attention due to the growing availability of historical competition data and the need for quantitative decision support in sports management. Among these tasks, Olympic medal prediction is a representative and challenging problem, as it involves complex temporal dependencies and is influenced by multiple socio-economic and sporting factors across different countries and Olympic cycles [4].

Early forecasting studies primarily relied on statistical models such as Autoregressive Integrated Moving Average (ARIMA), which model temporal correlations through linear autoregressive and moving-average component [5]. Although effective for stationary sequences, these models have limited capability in capturing nonlinear relationships commonly observed in real-world data. To address this issue, machine learning methods including decision trees, random forests, and gradient boosting algorithms have been introduced into prediction tasks [6]. In particular, Extreme Gradient Boosting (XGBoost) has demonstrated strong predictive performance by combining multiple weak learners and automatically identifying important features [7]. However, such approaches generally treat temporal observations as independent samples and may fail to model long-range sequential dependencies.

The emergence of deep learning has significantly advanced time-series forecasting research. Recurrent Neural Networks (RNNs) [8] and Long Short-Term Memory (LSTM) networks [9] were proposed to capture temporal patterns and long-term dependencies in sequential data. Subsequently, Bidirectional Long Short-Term Memory (BiLSTM) networks improved feature extraction by simultaneously learning forward and backward temporal information [10]. Nevertheless, recurrent architectures often encounter difficulties in identifying the most influential time steps and may exhibit degraded performance when dealing with long sequences.

To further enhance sequence modeling capability, attention mechanisms [11] and Transformer-based architectures have attracted increasing research interest. The Transformer architecture replaces recurrent operations with self-attention mechanisms, enabling global dependency modeling and efficient parallel computation [12]. Building upon this framework, numerous variants have been proposed for multivariate time-series forecasting, including Informer, Autoformer [13], and Improved Transformer (iTransformer). Although these models achieve promising results on large-scale datasets, they typically require abundant training samples and may suffer from instability or overfitting when applied to small-sample forecasting scenarios.

Despite substantial progress in forecasting research, several challenges remain in Olympic medal prediction [14]. First, medal sequences exhibit strong nonlinear fluctuations influenced by economic conditions, host-country effects, and sports development policies, making accurate temporal modeling difficult. Second, the importance of different features varies considerably across countries and Olympic cycles, while many existing methods lack adaptive feature weighting mechanisms. Third, critical Olympic years often have disproportionate impacts on future medal performance, yet conventional forecasting models may fail to adequately emphasize these key time steps. Finally, most existing studies adopt global temporal splitting strategies, which can introduce evaluation bias for countries with limited historical records and reduce model generalization capability.

To address these limitations, this study proposes AG-BiLSTM, a hybrid deep learning framework that integrates BiLSTM, multi-head self-attention, and gated feature fusion. The proposed model utilizes BiLSTM to capture bidirectional temporal dependencies, employs multi-head self-attention to identify influential Olympic cycles, and introduces a gated fusion mechanism to adaptively learn feature importance. Furthermore, a country-level temporal splitting strategy is designed to ensure a fairer evaluation process across countries with different data densities.

The major contributions of this work are summarized as follows:

(1) A novel AG-BiLSTM architecture is proposed by integrating BiLSTM, multi-head self-attention, and gated feature fusion, enabling simultaneous modeling of temporal dependencies and adaptive feature importance.

(2) A country-level independent temporal splitting strategy is introduced to alleviate evaluation bias caused by imbalanced historical records and improve model generalization.

(3) Extensive experiments on a self-constructed Olympic medal dataset demonstrate that the proposed model consistently outperforms representative statistical, machine learning, and deep learning baselines in gold, silver, and bronze medal prediction tasks.

2. Methods

This paper presents the AG-BiLSTM model for Olympic medal prediction. The proposed architecture combines a bidirectional LSTM for modeling temporal dependencies, a multi-head self-attention mechanism to weigh the importance of different cycles, and a gated fusion module that adaptively selects relevant features.

Given a country c and Olympic cycle t, the input feature vector is defined as

$$\mathbf{x}_t^c = [GDP_t^c, host_t^c, events_t^c, athletes_t^c, g_{t-1}^c, s_{t-1}^c, b_{t-1}^c]^\top \tag{1}$$

Here, GDP_t^cdenotes the normalized GDP feature, $host_t^c$denotes host status, $events_t^c$denotes the number of events, $athletes_t^c$denotes the number of key athletes, and g, s, and $b$denote historical medal counts. For a historical window of length T, the input sequence is

$$\mathbf{X}^c = [\mathbf{x}_{t-T+1}^c, \ldots, \mathbf{x}_t^c] \in \mathbb{R}^{T\times d} \tag{2}$$

where $d$is the feature dimension. For mini-batch training, the model input is $\mathcal{X} \in \mathbb{R}^{B\times T\times d}$, where $B$is the batch size. The prediction target is

$$\mathbf{y}_t^c = [g_t^c, s_t^c, b_t^c]^\top \tag{3}$$

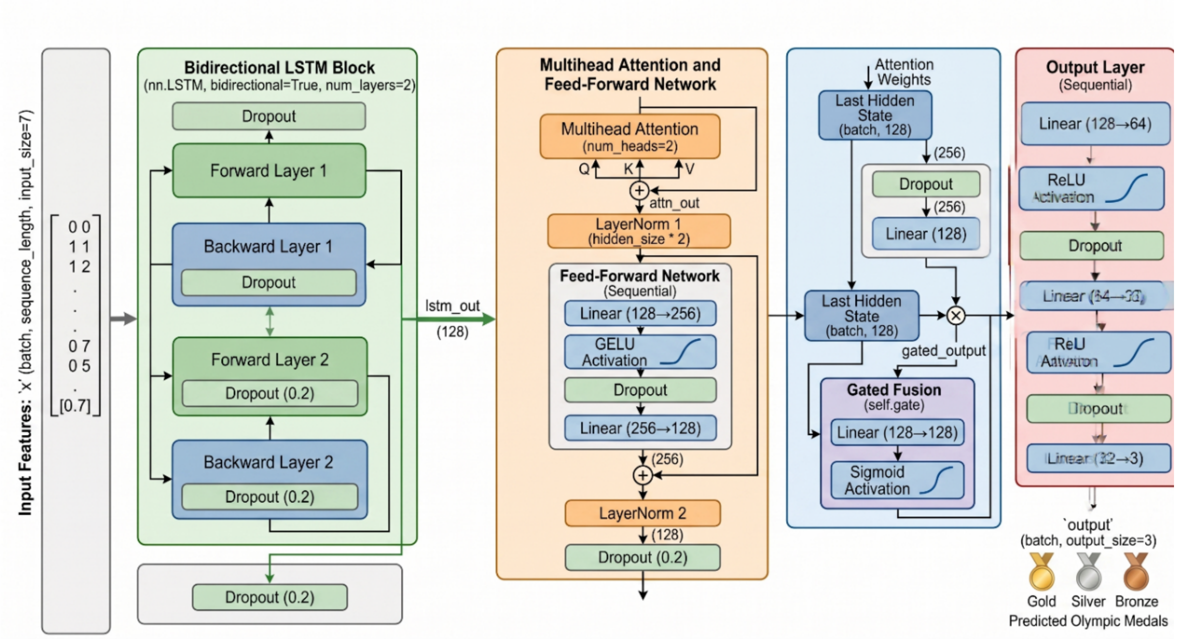
The learning objective is to estimate a nonlinear mapping

$$\hat{\mathbf{y}}_t^c = f_\theta(\mathbf{X}^c) \tag{4}$$

where $f_\theta$denotes AG-BiLSTM with learnable parameters θ. The input tensor is first sent to the BiLSTM module.

The overall architecture of AG-BiLSTM is illustrated in Figure 1. The model comprises seven components: an input module, a BiLSTM module, a multi-head self-attention mechanism module [15], a residual connection and layer normalization module, a feed-forward network module, a gated fusion module, and a multi-layer MLP output module.

Figure 1: The overall architecture of AG-BiLSTM



2.1 BiLSTM Module

The BiLSTM [16] module addresses the challenge of capturing both past and future temporal dependencies in Olympic performance sequences, which are inherently bidirectional. This module extracts bidirectional temporal features from the Olympic sequence data.

For each time step τ, a standard LSTM unit computes the following gate and state updates:

$$\mathbf{i}_\tau = \sigma(\mathbf{W}_{xi}\mathbf{x}_\tau + \mathbf{W}_{hi}\mathbf{h}_{\tau-1} + \mathbf{b}_i) \tag{5}$$

$$\mathbf{f}_\tau = \sigma(\mathbf{W}_{xf}\mathbf{x}_\tau + \mathbf{W}_{hf}\mathbf{h}_{\tau-1} + \mathbf{b}_f) \tag{6}$$

$$\mathbf{o}_\tau = \sigma(\mathbf{W}_{xo}\mathbf{x}_\tau + \mathbf{W}_{ho}\mathbf{h}_{\tau-1} + \mathbf{b}_o) \tag{7}$$

$$\tilde{\mathbf{c}}_\tau = \tanh\left(\mathbf{W}_{xc}\mathbf{x}_\tau + \mathbf{W}_{hc}\mathbf{h}_{\tau-1} + \mathbf{b}_c\right) \tag{8}$$

$$\mathbf{c}_\tau = \mathbf{f}_\tau \odot \mathbf{c}_{\tau-1} + \mathbf{i}_\tau \odot \tilde{\mathbf{c}}_\tau \tag{9}$$

$$\mathbf{h}_\tau = \mathbf{o}_\tau \odot \tanh\left(\mathbf{c}_\tau\right) \tag{10}$$

Here, $x_\tau$denotes the input feature vector at time step τ(e.g., normalized performance metrics such as athlete counts, GDP, or historical medal shares for a given Olympic cycle); $h_{\tau-1}$and $c_{\tau-1}$are the previous hidden state and cell state, respectively; W and b are learnable weight matrices and bias vectors for the input, forget, output, and candidate gates; $\sigma(\cdot)$is the sigmoid activation function; $\odot$represents element-wise multiplication; and $\tanh(\cdot)$is the hyperbolic tangent function. These equations enable the LSTM to selectively remember or forget information across time steps, mitigating vanishing gradient issues in long Olympic sequences.

The forward and backward hidden states are calculated as $\vec{h}_\tau$ and $\overleftarrow{h}_\tau$. They are concatenated to obtain the temporal representation:

$$\mathbf{h}_\tau = [\vec{\mathbf{h}}_\tau; \overleftarrow{\mathbf{h}}_\tau] \tag{11}$$

The complete BiLSTM output is $H = \{h_1, h_2, \dots, h_T\}$, where $T$is the total number of Olympic cycles in the sequence. This bidirectional representation enriches each time step with contextual information from both preceding and succeeding cycles. This output $H$is then passed directly to the Multi-Head Self-Attention Module to dynamically select critical Olympic cycles that are most relevant for medal prediction.

2.2 Multi-Head Self-Attention Module

While BiLSTM captures temporal dependencies, not all Olympic cycles contribute equally to future medal predictions (e.g., recent cycles or those with major policy shifts may be more informative). The attention module assigns different weights to different time steps, enabling the model to focus on the most salient historical patterns.

The BiLSTM output $H$is first projected into Query, Key, and Value matrices:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}^Q, \mathbf{K} = \mathbf{H}\mathbf{W}^K, \mathbf{V} = \mathbf{H}\mathbf{W}^V \tag{12}$$

For h attention heads, the dimension of each head is $d_k = d_{model}/h$. For the i-th head, the attention score, attention weight, and head output are:

$$\mathbf{A}^{(m)} = \mathrm{softmax}\left(\frac{\mathbf{Q}^{(m)}(\mathbf{K}^{(m)})^\top}{\sqrt{d_k}}\right)\mathbf{V}^{(m)} \tag{13}$$

the multi-head attention output is

$$\mathbf{H}_{\mathrm{att}} = \mathrm{Concat}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(M)})\mathbf{W}^O \tag{14}$$

In these equations, $\boldsymbol{W}^Q, \boldsymbol{W}^K, \boldsymbol{W}^V, \boldsymbol{W}^O$ are learnable projection matrices; $d_k$is the scaling factor to prevent vanishing gradients in softmax; and $Q_i, K_i, V_i$are the projections for the i-th head. This mechanism computes similarity scores between time steps and produces a weighted aggregation of values, highlighting influential Olympic cycles. The output of the multi-head attention, combined with the original BiLSTM representation, proceeds to the Residual Normalization and Feed-Forward Module for stabilized feature refinement.

2.3 Residual Normalization and Feed-Forward Module

Deep architectures can suffer from gradient issues or dimensional mismatches between sub-modules. Since the BiLSTM output and attention output may have different dimensions, the BiLSTM output is first projected. The first residual connection and layer normalization are defined as:

$$\mathbf{H}' = \mathrm{LayerNorm}(\mathbf{H}_{\mathrm{proj}} + \mathbf{H}_{\mathrm{att}}) \tag{15}$$

for a feature vector $\mathbf{z}$, layer normalization is

$$\mathrm{LayerNorm}(\mathbf{z}) = \gamma \odot \frac{\mathbf{z}-\mu}{\sqrt{\sigma^2+\epsilon}} + \beta \tag{16}$$

where μ and σ are the mean and standard deviation computed across features, and $\gamma, \beta$are learnable scale and shift parameters.

The feed-forward network then performs nonlinear feature transformation, followed by a second residual connection. The output $\mathbf{H}''$ is a stable temporal representation and is sent to the gated fusion module.

2.4 Gated Fusion Module

The gated fusion module learns feature-level importance weights. The final time-step representation is selected as $\mathbf{h}_T''$. The gate vector is computed by

$$\mathbf{g} = \sigma(\mathbf{W}_g \mathbf{h}_T'' + \mathbf{b}_g) \tag{17}$$

The gated representation is

$$\mathbf{z} = \mathbf{g} \odot \mathbf{h}_T'' \tag{18}$$

Here, g acts as a soft gate that controls information flow, learned via sigmoid. This allows the model to emphasize the most predictive aspects of Olympic history for each country. The gated representation z is forwarded to the MLP Output Module for final medal prediction.

2.5 MLP Output Module and Overall Computational Pipeline

The MLP output module maps the gated representation to the three medal targets:

$$\hat{\mathbf{y}} = \mathbf{W}_o \cdot \mathrm{ReLU}(\mathbf{W}_h \mathbf{z} + \mathbf{b}_h) + \mathbf{b}_o \tag{19}$$

where W_h and b_h are the weights and bias of the hidden layer, W_o and b_o are those of the output layer, and $\hat{y}$denotes the predicted vector of gold, silver, and bronze medals.

The overall computational pipeline is:

$$H_{BiLSTM} = BiLSTM(X) \tag{20}$$

$$H_{Att} = MultiHeadAttention(H_{BiLSTM}) \tag{21}$$

$$H_{Res1} = LayerNorm(H_{BiLSTM} W_R + H_{Att}) \tag{22}$$

$$H_{Res2} = LayerNorm(H_{Res1} + FFN(H_{Res1})) \tag{23}$$

$$h_{gate} = Gate(H_{Res2}) \tag{24}$$

$$\hat{Y} = MLP(h_{gate}) \tag{25}$$

During training, the prediction is compared with the true medal vector by mean squared error:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \parallel \hat{\mathbf{y}}_i - \mathbf{y}_i \parallel^2 \tag{26}$$

where N is the number of samples, $\hat{y}_i$ is the predicted medal vector (gold, silver, and bronze), and y_i is the corresponding ground-truth medal vector for the i-th country.

Here, the summation runs over gold, silver, and bronze medals. Through this pipeline, AG-BiLSTM connects temporal modeling, critical time-step selection, adaptive feature weighting, and final medal prediction in a single computable framework.

3. Experiments

This section evaluates the AG-BiLSTM model's performance on a dataset used for Olympic medal prediction, demonstrating the model's reliability and robustness.

3.1 Experimental Data

3.1.1 Dataset Overview

The experiment uses data from 19 major Olympic-participating countries across 16 Olympic editions from 1964 to 2024. The dataset consists of medal records and influencing features, covering three medal types.

Medal Data: Medal records were obtained from the official statistical reports of the International Olympic Committee (IOC), ensuring reliability and authority. Due to political events, medal data for some countries in 1980 were missing; therefore, linear interpolation was used to estimate the missing values:

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} * (x - x_1) \tag{27}$$

where (x_1, y_1) and (x_2, y_2) denote two known coordinate points on the straight line, x denotes the independent variable of the point to be solved, and y denotes the corresponding dependent variable to be calculated via linear interpolation.

This completed the 1980 medal records for the affected countries.

Influencing Features: The selected features include GDP, host status, number of events, and key players. GDP refers to the sum of a country's GDP over the four years preceding each Olympic year, normalized to the range [0,1]. Host status is coded as 1 for the host country and 0 otherwise. Number of Events represents the number of sporting events in which a country participates, while Key Player denotes the number of star athletes in the country. These datas were obtained from the official statistical reports of the IOC and International Monetary Fund(IMF).

The data samples adopted in this experiment are shown in Table 1:

Table 1: Data Samples

Year	NOC	Gold	Silver	Bronze	Total	GDP	Host	Number of events	Key Player
2024	United States	40	44	42	126	1	0	329	14
2024	China	40	27	24	91	0.717487	0	329	16
2024	Japan	20	12	13	45	0.190809	0	329	9
2024	Australia	18	19	16	53	0.064974	0	329	7
2024	France	16	26	22	64	1	1	329	6
2024	Netherlands	15	7	12	34	0.04287	0	329	5
2024	Great Britain	14	22	29	65	0.128611	0	329	10

3.1.2 Dataset Splitting

Unlike traditional global temporal splitting, this study employs a country-level independent splitting strategy. Data from each country is divided chronologically into a training set (80%), validation set (15%), and test set (5%) in an 8:1.5:0.5 ratio, ensuring that each country contributes to the test set. The training, validation, and test samples from all countries are subsequently merged, yielding a final training set of 520 samples, a validation set of 98 samples, and a test set of 33 samples, covering all 19 countries [17].

3.2 Experimental Setup

All experiments were conducted under identical hardware and software environments. The hardware configuration comprised an AMD Ryzen 9 8940HX CPU, 32 GB of RAM, and an NVIDIA RTX 5060 discrete GPU. Experiments were executed on PyTorch 2.0, with Python 3.9 and CUDA 11.7 for GPU-accelerated training.

3.3 Baseline Models and Evaluation Metrics

To evaluate the effectiveness of AG-BiLSTM, several representative medal prediction and time-series forecasting models were selected as baselines. ARIMA is used as a traditional time-series benchmark for modeling linear autoregressive and moving-average patterns. XGBoost provides strong nonlinear fitting and feature mining capabilities and is widely used in prediction and regression tasks. LSTM captures long-term temporal dependencies, while BiLSTM further enhances temporal representation by learning dependencies in both forward and backward directions. Transformer employs global attention to model long-range sequence dependencies, and iTransformer improves multivariate time-series forecasting through enhanced feature mapping and attention computation.

Model performance is evaluated from three perspectives: predictive accuracy, average error, and goodness of fit. The adopted metrics are defined as follows.

Root Mean Square Error (RMSE): RMSE measures the overall deviation between predicted and actual medal counts, with smaller values indicating better predictive accuracy. The formula is:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{28}$$

where y_i denotes the true value, $\hat{y}_i$ denotes the predicted value and m denotes the number of samples.

3.4 Main Results

The prediction results of AG-BiLSTM are presented in the following tables. Table 2 reports the average RMSE comparison for a subset of countries across all evaluated models. Among the 19 sample countries, 11 achieved higher predictive accuracy under the AG-BiLSTM model than under any other model. Table 3 presents the gold, silver, bronze, and average RMSE values for each model. The average RMSE for medal prediction is lower than that of all other models: compared to the traditional ARIMA model, the improvement is 51%, likely attributable to AG-BiLSTM's capacity to accept multi-feature inputs; compared to the machine learning model XGBoost, the improvement is 12%, potentially due to AG-BiLSTM's integration of an attention mechanism that enables the model to focus on critical time steps such as host country years; compared to the deep learning models LSTM and BiLSTM, the improvements are 9% and 1%, respectively, possibly because AG-BiLSTM augments the BiLSTM backbone with a gated feature fusion module capable of automatically selecting important features. It is further noteworthy that in gold medal prediction, the model substantially outperforms all other models, with improvements of 52%, 23%, 11%,18%, 29%, and 10% over ARIMA, LSTM, BiLSTM, Transformer, iTransformer, and XGBoost, respectively.

Table 2: Per-Country Average RMSE Comparison

Model	Brazil	Sweden	Hungary	Spain	Norway	Uzbekistan
AG-BiLSTM	1.64	1.95	1.70	2.05	1.83	2.49
ARIMA	1.77	2.06	15.06	3.43	3.14	3.07
LSTM	2.59	1.59	1.65	2.41	1.43	2.53
BiLSTM	1.87	1.55	1.67	2.54	1.27	3.50
Transformer	2.64	1.69	2.22	2.80	1.83	2.51
iTransformer	4.36	1.51	1.16	2.72	1.43	2.87
XGBoost	3.37	1.59	1.72	2.75	1.53	2.77

Table 3: RMSE Comparison by Medal Type Across Models

Model	Gold RMSE	Silver RMSE	Bronze RMSE	Average RMSE
AG-BiLSTM	3.75	3.61	3.87	3.74
ARIMA	7.87	7.54	7.45	7.62
LSTM	4.89	3.48	3.49	3.95
BiLSTM	4.23	3.52	3.62	3.79
Transformer	4.55	4.12	3.53	4.06
iTransformer	5.26	4.41	4.81	4.82
XGBoost	4.15	4.37	4.32	4.28

3.5 Visualization Analysis

To better illustrate which medal types are predicted most accurately by the model under small-sample conditions and for which countries, the countries with the lowest RMSE values for gold, silver, and bronze medal predictions are presented separately. As shown in the Figure 2, the model achieves the best performance in gold medal prediction for Sweden and Italy; As shown in the Figure 3 for silver medal prediction, Brazil yields the best results; and as shown in the Figure 4 for bronze medal prediction, Norway, Brazil, and Hungary all demonstrate excellent predictive performance.

Figure 2: The countries with the lowest RMSE values for Gold Medal

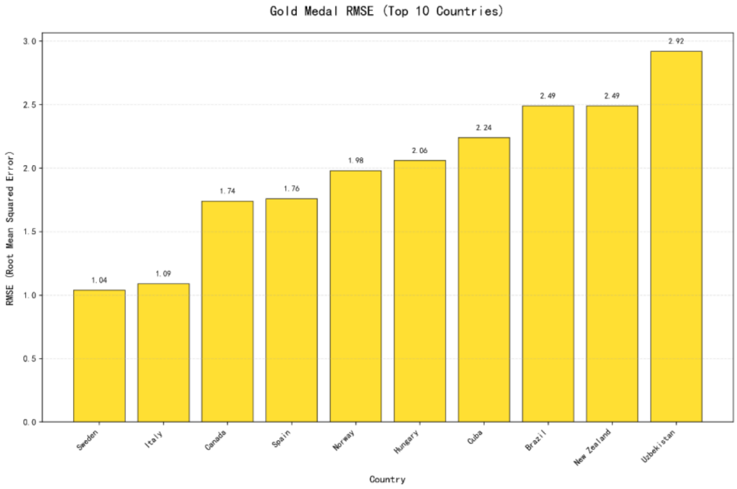


Figure 3: The countries with the lowest RMSE values for Silver Medal

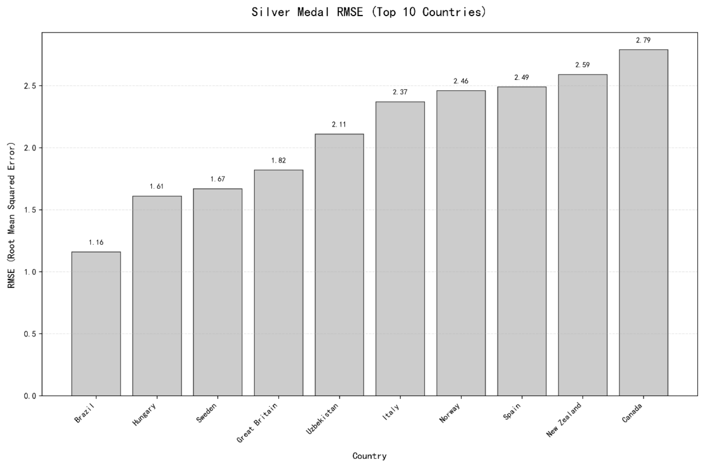
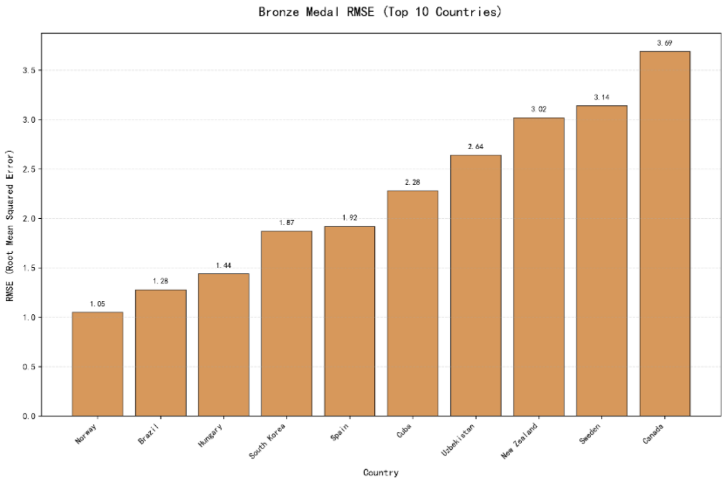


Figure 4: The countries with the lowest RMSE values for Bronze Medal



4. Conclusion

This study addresses key challenges in Olympic medal prediction, including capturing nonlinear temporal dependencies, emphasizing critical features, and improving predictive accuracy for countries with limited data. The AG-BiLSTM model significantly enhances both representational capacity and generalization performance under small-sample conditions. Extensive experimental results demonstrate that AG-BiLSTM outperforms existing machine learning and deep learning models on core metrics, including RMSE.

While the model performs well, it has certain limitations. The dataset only includes 19 major countries, and future research could extend this to include more countries and longer time periods. Additionally, the model relies primarily on historical statistical features, excluding micro-level factors such as individual athlete performance. Future work could integrate graph neural networks to model inter-country relationships and multimodal data to further enhance predictive accuracy and robustness.

References

[1] Zhang, Y., & Yan, J. (2023). Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. ICLR 2023.

[2] Sezer, O., Gudelek, M., & Ozbayoglu, A. (2020). Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019. Applied Soft Computing, 90.

[3] Lana, I., Ser, J., Velez, M., & Vlahogianni, E. (2018). Road Traffic Forecasting: Recent Advances and New Challenges. IEEE Intelligent Transportation Systems Magazine, 10(2), 93-109.

[4] Sayeed, R., Hassan, M., Rahman, M., Zaman, F., Ahmed, S., & Miah, M. (2025). Machine Learning Models for Predicting Olympic Medal Outcomes. 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 3, 1-6.

[5] Siami-Namini, S., Tavakoli, N., & Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine Learning and Applications (icmla), 1394-1401.

[6] Qiu, X., Zhang, L., Suganthan, P., & Amaratunga, G. (2017). Oblique Random Forest Ensemble Via Least Square Estimation for Time Series Forecasting. Information Sciences, 420, 249-262.

[7] Cai, R., Xie, S., Wang, B., Yang, R., Xu, D., & He, Y. (2020). Wind Speed Forecasting Based on Extreme Gradient Boosting. IEEE Access, 8, 175063-175069.

[8] Torres, J., Hadjout, D., Sebaa, A., Martinez-Alvarez, F., & Troncoso, A. (2020). Deep Learning for Time Series Forecasting: A Survey. Big Data, 9(1), 3-21.

[9] Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. IEEE Communications Magazine, 57(6), 114-119.

[10] Du, S., Li, T., Yang, Y., & Horng, S. (2020). Multivariate Time Series Forecasting Via Attention-Based Encoder–decoder Framework. Neurocomputing, 388, 269-279.

[11] Shih, S., Sun, F., & Lee, H. (2019). Temporal Pattern Attention for Multivariate Time Series Forecasting. Machine Learning, 108(8), 1421-1441.

[12] Khan, S., Naseer, M., Hayat, M., Zamir, S., Khan, F., & Shah, M. (2022). Transformers in Vision: A Survey. ACM Computing Surveys, 54(10S).

[13] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. NEURIPS 2021.

[14] Wang, Z., Yan, Y., & Wang, L. (2025). Olympic Medal Prediction Via Hybrid Time-Series and Static Feature Modeling. 2025 10th International Conference on Computer and Communication System, Icccs, 25-30.

[15] Luo, L., Yang, Z., Yang, P., Zhang, Y., Wang, L., Lin, H., & Wang, J. (2018). An Attention-Based BiLSTM-CRF Approach to Document-Level Chemical Named Entity Recognition. Bioinformatics (Oxford, England), 34(8), 1381-1388.

[16] Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment Analysis of Comment Texts Based on BiLSTM. IEEE Access, 7, 51522-51532.

[17] Bergmeir, C., & Benítez, J. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, (1), 192-213.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.